TextAnalyst 2

User Manual

Copyright Notices

Trademarks

Table of Contents

CHAPTER 6: WORKING WITH A TEXT BASE 50

CHAPTER 7: CUSTOMIZING TEXTANALYST 53

CHAPTER 8: USING DICTIONARY EDITOR 56

Chapter 1: About this Manual

This section covers:

- Manual Organization
- Manual Conventions
- Definitions

Manual Organization

This manual is divided into eight chapters.

Chapter 1, *About this Manual*, provides an overview of the manual's organization, describes the conventions we use, and lists key definitions that can help speed the user's understanding of TextAnalyst.

Chapter 2, *Introduction to Natural Language Text Analysis*, helps you understand the basics of Natural Language Text Analysis and TextAnalyst's capabilities. This chapter also provides a list of new features in TextAnalyst 2.1 and a description of the TextAnalyst COM.

Chapter 3, *Installation*, covers customer support provisions, the Megaputer license agreement and copyright, and system requirements. Chapter 3 then describes the installation procedures for TextAnalyst 2.1.

Chapter 4, *The TextAnalyst Interface*, discusses basic navigation, which includes startup and opening the sample text. This chapter also covers accepted file formats and organization of the knowledge base. The details of the TextAnalyst workspace are available in this chapter as well.

Chapter 5, *Working with a Single Text*, tells you how to create a knowledge base and understand the text analysis. It covers the tabs in the View pane, navigation and editing of the knowledge base, and text summarization. Chapter 5 also covers semantic search, standard search, and exporting to .CSV and .HTML file formats.

Chapter 6, *Working with Several Texts*, tells you how to manage analyses, summarizations, and searches that deal with more than one text at a time.

Chapter 7, *Customizing TextAnalyst*, helps you use the General Settings and View Settings selections to tailor TextAnalyst to suit your needs.

Chapter 8, *Using Dictionary Editor*, helps you understand the settings in the dictionary editor and, if you choose, use them to create and save a custom dictionary.

Manual Conventions

This manual uses the following text characteristics to convey special meaning.

- **Boldface Type** indicates an on-screen element, such as a button, a menu, or a toolbar icon. We use boldface type to show most items you need to locate, select, or click.

- UPPERCASE TYPE indicates a file name or an operating system command.

- The " | " symbol separates a sequence of menu commands. For instance, if you see, "Choose **File | Create**," you should:

  1. Go to the Main menu and click **File** on the menu bar at the top of the screen. The File drop-down menu appears.

  2. Move the cursor to **Create** on the dropdown menu and left-click the mouse.

Definitions

The following definitions can help you understand TextAnalyst more quickly.

- **Concept** – Refers to a word or words (term or terms) TextAnalyst identifies as significant in your text. Concepts appear as hyperlinks in text and as list items in tree structures.

- **Text** – Refers to a document you load in TextAnalyst. Both .TXT and RTF file formats are acceptable.

- **Semantic network** – A tree structure of concepts from your text and the relationships between them. This is a concise representation of your text.

- **Knowledge base** – The collection of your text, the semantic network related to your text, any edits you made, the results of your analyses, and hyperlinks within your text.

- **Semantic search** – Semantic search is synonymous with Natural Language Query. You type a question in conventional, common English, and TextAnalyst returns results for your examination.

- **Semantic weight** – The semantic weight of a concept is a measure of its importance in your document. This is the number closest to a concept in a tree structure when measuring semantic weight. The semantic weight of the relationship between a concept and its parent concept is the leftmost number in a pair when measuring semantic weight. This number shows the measure of the strength of the relationship between the concept and its parent.

Chapter 2: Introduction to Natural Language Text Analysis

This section covers:

- About Natural Language Text Analysis
- TextAnalyst Algorithms
- What TextAnalyst Can Do
- What's New in TextAnalyst 2.1
- TextAnalyst COM

About Natural Language Text Analysis

This section briefly outlines the development of text mining, as well as TextAnalyst's origins. The purpose of this section is to give you a clear understanding of how TextAnalyst works and why it is successful.

Computer-Human Communication and Neural Network Technology

Natural language text analysis developed with the goal of rapid and efficient navigation through large volumes of textual information. Numerous professions require such activities, and benefit from automation where possible.

Computer-human communication, a branch of Artificial Intelligence (AI), uses linguistics to analyze text. Linguistic rules lead to a cognitively organized text representation object. The result of the analysis is a semantic network. The drawback of this method is that human intervention and reprogramming is required each time the subject matter of the analysis changes. Part of the goal is accomplished, but human effort is still required.

Another significant contributor to natural language text analysis is artificial neural network technology. This is a method of parallel processing designed to process information similar to the way neurons of the human brain do. Neural networks are capable of self-learning, so they have a potential of analyzing text automatically and producing a concise representation of the meaning of the text on the output. Human guidance is possible, but not required in this technique. The homogeneous processing mechanism in this approach has a lot of promise. Unfortunately, however, known neural network architectures are often unsuited for successful analytical processing of natural language texts.

Best of Both Worlds: Megaputer's TextAnalyst

TextAnalyst processes text automatically as neural networks, but produces semantic structures as an end product. TextAnalyst views the original text as a sequence of symbols, taking snapshots of groups of symbols at a time. A hierarchical neural network results, with frequency of occurrence being most important. The relationship among symbols also affects the hierarchy. TextAnalyst ignores common words, such as the articles "a", "and", and "the".

By creating a neural network of symbols and producing a semantic network, TextAnalyst creates a knowledge base useful for, but not limited to, the following:

- Topic structure display

- Distillation of the meaning of the text

- Text summarization of user-specified length

- Semantic information retrieval

- Word search and semantic search

- Clustering

The software requires human intervention only if text in a language other than English were loaded. In this case, the TextAnalyst would require input regarding the articles and other common words it should ignore.

Megaputer provides you with a state-of-the-art tool for natural language text analysis. We wish you success with TextAnalyst 2.1.

TextAnalyst Algorithms

This section details the process TextAnalyst uses to prepare your text for analysis. Completion of this process results in an established semantic network.

Preprocessing

TextAnalyst begins analysis of your text by preprocessing. This preliminary step involves elimination of supplementary words and identification of word stems.

Elimination of Supplementary Words

Supplementary, or common, words do not enter into the semantic analysis. During preprocessing, these words are eliminated because they carry no semantic meaning. Articles of speech, such as a, an, and the, represent the kind of words that preprocessing eliminates. Preprocessing allows TextAnalyst to filter out meaningless elements.

Identification of Word Stems

TextAnalyst identifies word stems and separates these stems from their accompanying prefixes, suffixes, and endings (morphemes). The result is an analysis of the stems alone, which provides a clearer picture of their frequency and relationship to other words in the text. While performing the analysis with individual stems, the network still holds information about complete words.

Language-Specific Programming

Preprocessing is the only place phase where language dependency enters into the TextAnalyst text mining technique. All other components of this technology are language independent and do not require human intervention.

Because elimination of supplementary words and identification of word stems are both language-specific activities, TextAnalyst requires programming of these elements to ensure recognition. TextAnalyst requires programming that is specific to the language of the text it will process.

Statistical Analysis

After preprocessing is complete, the neural network holds all important words and word combinations for your text, as well as frequency of occurrence. TextAnalyst assesses frequencies of joint occurrences among semantic elements. A graph-like structure results with statistical weights of both words in the nodes and words having joint occurrences.

TextAnalyst establishes individual statistical weights of the words and statistical weights of their relationships to each other.

Renormalization

During renormalization, TextAnalyst adjusts the statistical weights of the words with respect to the strength of their relationship with other words in the sentence. The neural

network evolves into a stable configuration corresponding to the minimum of an energy function characterizing the network.

Renormalized weights of the words and the relationships between them are called semantic weights. The resulting reshaped, graph-like structure is called a semantic network. This network provides the most important words and word combinations from your text and the relationships between them. Words and word combinations in the semantic network are called semantic concepts.

Following renormalization, the semantic network is ready for use in text mining.

What TextAnalyst Can Do

TextAnalyst2.1 provides text-mining capabilities that can be useful to you in doing the following. This list is not exhaustive.

- **Semantic analysis** – By default, TextAnalyst automatically performs a semantic analysis of your text when you load it. The result is an expandable and collapsible tree structure of concepts from your text and the relationships between them. This is a semantic network, a concise representation of your text.

- **Navigation** – TextAnalyst links key concepts in your text to concepts in your analyses. Since these concepts are also in the semantic network, you can navigate through the semantic network and hyperlinks between an analysis and the original text.

- **Summarization** – You can generate a summary containing the most important sentences in your text. TextAnalyst also lets you adjust the number of sentences to a few key sentences or a broader overview.

- **Natural language queries** – You can search for information by typing a question for TextAnalyst in conventional written English.

- **Knowledge base development** – TextAnalyst keeps a knowledge base for you containing your text(s), semantic network, edits, results of analyses, hyperlinks, and any related dictionaries.

- **Topic structure organization** – You can view the semantic network of concepts by topic structure for an overview of topic organization.

- **Organization into clusters** – Whether you load one topic or several, TextAnalyst provides a view of clustered topics you can explore in the tree structure.

- **Dictionary development** – You can add to and delete from the TextAnalyst dictionary, and also indicate important words for the analysis.

- **Focused analysis** – By narrowing the criteria in a search, analysis, or summary, you can pinpoint the focus of the text mining.

What's New in TextAnalyst 2

Version 2 of TextAnalyst contains the following new features:

- The user interface has a new look, providing a more efficient work environment. You now can easily switch between different views of the knowledge base, and you can customize the appearance of the workplace to your liking.

- New user-defined settings and controls enhance the quality and flexibility of system performance.

- The reworked analytical algorithms are now capable of robust processing, analyzing up to 50 Mbytes of text in a single session.

- Component Object Model-based (COM-based) internal architecture now makes TextAnalyst easily extendable and upgradeable.

- You can export a knowledge base to .HTML file format for navigating in a Web browser. This allows you to share results of your analyses over the Web.

- You can access a customizable list of unconditionally included or excluded words.

- You can access a customizable list of synonyms for more accurate recognition of the terms of interest.

- An improved mechanism for semantic search gives you increased control over the volume of retrieved results.

- Improved capabilities are available for communicating with external applications and for representing the results of exploration.

- TextAnalyst now processes text in .RTF file format.

- You can export the significant concepts of your text and their statistical relationships to Microsoft Excel from the semantic network TextAnalyst automatically creates.

- Individual modules of TextAnalyst are available as COM components for a simple integration in external analytical applications.

TextAnalyst COM

A version of TextAnalyst is available implementing the specification for Microsoft's Component Object Model (COM), also known as ActiveX. The specification establishes a standard for communication between software components. The exploration engines are contained in dynamic link libraries.

ActiveX controls can be accessed by applications other than TextAnalyst. As a result, you can incorporate the semantic analyses algorithms of TextAnalyst into other applications, such as browsers, e-mail services, search systems, relational database management systems (RDBMS), and so on. This gives you the power and versatility of TextAnalyst's analytical algorithms built into your favorite software product or your company's own products and operational systems.

For more information about TextAnalyst COM, please either contact your Megaputer distributor or e-mail Megaputer at tmcomponents@megaputer.com.

Chapter 3: Installation

This section covers:

- Customer Support

- Megaputer License Agreement and Copyright

- System Requirements

- TextAnalyst 2.1 Installation

Customer Support

Free Standard Customer Support

The license fee covers the following support services:

- Initial telephone consultations on text mining with TextAnalyst for three hours or for the duration of one project

- Telephone and e-mail technical support for one month

- An 80% discount on any updates of TextAnalyst released within three months of the TextAnalyst 2 delivery date

Full Customer Support Contract

One- and two-year Full Customer Support Contracts are available. These include:

- Free software upgrades within the current version

- An 80% discount on software upgrades to the next version

- Access to the telephone technical support hotline and e-mail technical support

- Participation in user seminars at no charge

- Information about new products and events

Rates for a Full Customer Support Contract are:

For:	The fee is:
1 year	18% of the corresponding license fee
2 years	34% of the corresponding license fee

Megaputer License Agreement and Copyright

Megaputer License Agreement

CAREFULLY READ THE FOLLOWING LICENSE AGREEMENT PRIOR TO OPENING ANY OF THE DISK PACKS. BY OPENING A DISK PACK YOU INDICATE YOUR ACCEPTANCE OF THE TERMS OF THIS AGREEMENT.

Definitions

The following definitions apply to the terms as they appear in the license agreement:

Agreement means the present license agreement.

Megaputer means Megaputer Intelligence Inc.

Software means the computer program contained in this package, and any and all updates to this computer program received later.

Documentation means the user manual and all other printed materials accompanying the software.

Product means the software and documentation together.

Copyright/Proprietary Protection

The product is owned by Megaputer or its suppliers and is protected by international copyright laws and international trade provisions. You must treat this product exactly as any other copyrighted material. The agreement is considered breached and your right to use the product terminates automatically if you violate any provision of this agreement. In the event of license termination, you must immediately destroy all copies of the product or return them to Megaputer.

User Rights and Responsibilities

Megaputer grants you a nonexclusive license to:

Use a single copy of the software at a single location.

Make one copy of the software as a backup, or copy the software onto the hard disk of your computer and retain the original as a backup.

You may not:

Copy the documentation.

Modify, translate, or merge the software with another program, except for your personal use on a single computer. Any modifications to the software are subject to this agreement.

Disassemble, decompile, or make any attempt to discover the source code of the software.

Sublicense, rent, or lease any portion of the product.

Disclose to a third party the terms of your acquisition of the software.

Use an obsolete version of the software for which you received an upgraded or updated copy under this agreement. Upon upgrading or updating the software, you must delete the old copy from the computer. You must physically destroy the original disks or return them to Megaputer.

Limited Warranty
Megaputer warrants the installation disks on which the software is distributed to be free from defects in materials and workmanship and the software to perform substantially in accordance with the documentation for a period of 90 days from your receipt of the product.

If the product fails to comply with the warranty set forth above, Megaputer's entire liability and your exclusive remedy will be the replacement of the installation disks or, at Megaputer's option, Megaputer's reasonable effort to make the product meet the warranty set forth above. This limited warranty applies only if you return all copies of the product, along with the copy of your paid invoice, to Megaputer within 90 days from the date you received the product. Any replacing software will be warranted for the remainder of the original 90-day warranty period.

No Liability for Consequential Damages

In no event shall Megaputer or its suppliers be liable for any damages whatsoever (including, without limitation, damages for loss of profit, business interruption, loss of information, or any other particular loss) that result from the use of, or the inability to use, this Megaputer product, even if Megaputer has been advised of the possibility of such damages.

General

Megaputer product support is available only to registered users of the product.

This agreement constitutes the entire agreement between you and Megaputer and supersedes any prior agreements concerning the contents of this package. It shall not be modified except by a written consent of both parties dated later than the date of this agreement and signed by an authorized Megaputer representative.

System Requirements

This section provides hardware and software requirements for TextAnalyst 2.1.

Hardware

TextAnalyst 2.1 has the following minimal and optimal hardware requirements:

Minimal Requirements

- Intel Pentium I processor or better

- Windows 95/98/NT/2000/XP with 32 Mbytes RAM

- 30 Mbytes hard disk space

- Mouse

Optimal TextAnalyst Requirements

- Pentium III processor or better

- Windows 95/98/NT/2000/XP with 128 Mbytes RAM30 Mbytes hard disk space

- SVGA video card

By increasing processing power and RAM, you can reduce the processing time of the semantic analyses, while increasing the ability of TextAnalyst to work with large texts.

Software

The software requirement for TextAnalyst 2.1 is that you use one of the following three platforms:

- Microsoft Windows NT 4.0

- Windows 95/98

- Windows 2000

- Windows XP

TextAnalyst 2 Installation

This section provides instructions for both current and new TextAnalyst users. If you are a current user, we highly recommend that you uninstall your version(s) of TextAnalyst before installing version 2.

Uninstall Steps for Current Users

If you now use a version of TextAnalyst earlier than 2, follow these steps to uninstall your current version and any other versions of TextAnalyst. If this is your first version of TextAnalyst, skip ahead to the next section of this manual and begin installation.

To uninstall a previous version of this software:

1. From the Windows **Start** menu, choose **Settings | Control Panel**.

2. Double-click **Add/Remove Programs** from your Windows **Control Panel**.

3. From the alphabetized list of programs, select the version of TextAnalyst you are uninstalling.

4. Click the **Add/Remove** button, then click **Yes** in the **Confirm File Deletion** window. The **Remove Programs from Your Computer** window appears and indicates when uninstall is complete.

5. Click **OK**.

6. Repeat the above steps for all versions of TextAnalyst on your system.

You are now ready to begin installing TextAnalyst 2.

Installation Procedure for Both Current and New Users

Once a current user uninstalls all earlier versions of TextAnalyst, the installation procedure for current and new users is the same. You can install TextAnalyst either from the TextAnalyst 2.1 CD-ROM, or from the Megaputer Web site.

Installation from CD-ROM

To install TextAnalyst 2.1 from the CD-ROM, follow these steps:

1. Insert the **TextAnalyst 2.1 CD-ROM** into the CD-ROM drive.
   The TextAnalyst window should display automatically, allowing you to go to the next step of this procedure.
   If you have auto run disabled, no screen displays when you insert the TextAnalyst CD-ROM. If you do not see the TextAnalyst screen, do one of the following:

   - Start **Windows Explorer** and double-click the **SETUP.EXE** program in the **X:\TEXTANALYST\TEXTANALYST 2.1** folder. **X** is the drive letter of your CD-ROM drive.

   - Choose **Run** from the Windows **Start** menu and either browse for **SETUP.EXE** on your CD-ROM, or enter **X:\TEXTANALYST\TEXTANALYST 2.1\SETUP.EXE. X** is the drive letter of your CD-ROM drive.

The **TextAnalyst 2.1 Setup** screen should now appear on your screen.

2. Click the **Next** button to display the TextAnalyst license agreement.

3. Read the license agreement, and if you agree, click **Yes**.
    The **TextAnalyst 2.1 for MS Windows 95/98/NT README file** appears.

4. After reading the README file, click **Next** to continue the installation.
    The **Choose Destination Location** window appears, so you can choose an installation directory.

5. Decide where you wish to install the TextAnalyst 2.1 program files.
    By default, TextAnalyst 2.1 installs in **C:\PROGRAM FILES\MEGAPUTER INTELLIGENCE\MICROSYSTEMS\TEXTANALYST 2.1**.

6. If the default installation directory is acceptable, click **Next**. Otherwise, click **Browse** and indicate another directory for installation.

7. Click **Next**.
    The **Select Program Folder** window appears. Your program icons go in the folder you select.

8. Indicate a different program folder or accept the default, which is **TextAnalyst 2.1**.

9. Click **Next**.
    TextAnalyst 2.1 Setup copies and registers all TextAnalyst 2.1 files to your hard drive.

10. Click **Finish**.

11. When setup completes, restart your computer.
    You are now ready to run TextAnalyst 2.1.

Installation from the Megaputer Web Site

To install TextAnalyst 2.1 from the Megaputer Web site, you first load the software to obtain your registration number, then you purchase the software to receive your activation code.

Loading the Software
1. From your Web browser, go to **www.megaputer.com**.

2. Click **Try & Buy**, then click **Try**.
    You will see a table under a header that reads, **Try Megaputer software hands-on!**

3. In the "try" table, go to TextAnalyst 2.1 and click the checkboxes for **Try** and **Request User Manual** in the rightmost columns.

4. In the last row of the table, click the **Download selected** button and follow the prompts.

5. After download is complete, start **Windows Explorer** and go to the location in which you downloaded TextAnalyst.

6. Double-click the **ta2eva.exe** program.
    The **TextAnalyst 2.1 Setup** screen should now appear on your screen.

7. Click the **Next** button to display the TextAnalyst license agreement.

8. Read the license agreement, and if you agree, click **Yes**.
   The **TextAnalyst 2.1 for MS Windows 95/98/NT README file** appears.

9. After reading the README file, click **Next** to continue the installation.
   The **Choose Destination Location** window appears, so you can choose an installation directory.

10. Decide where you wish to install the TextAnalyst 2.1 program files.
    By default, TextAnalyst 2.1 installs in **C:\PROGRAM FILES\MEGAPUTER INTELLIGENCE\MICROSYSTEMS\TEXTANALYST 2.1**.

11. If the default installation directory is acceptable, click **Next**. Otherwise, click **Browse** and indicate another directory for installation.

12. Click **Next**.
    The **Select Program Folder** window appears. Your program icons go in the folder you select.

13. Indicate a different program folder or accept the default, which is **TextAnalyst 2.1**.

14. Click **Next**.
    TextAnalyst 2.1 Setup copies and registers all TextAnalyst 2.1 files to your hard drive.

15. Click **Finish**.

12. When setup completes, restart your computer.

Purchasing the Software
Now that TextAnalyst is loaded, you are able to use the following steps to give Megaputer your registration number and receive your activation code.

1. Start TextAnalyst as directed in chapter 4 under Basic Navigation.
   You are now ready to locate your registration number.

2. From the **Main** menu, click **File | Register**.

3. In the **Registration** dialog, copy your registration number from the top dialog box.

4. Go to the Megaputer Web site at www.megaputer.com.

5. Click **Try & Buy**, then click **Buy**.
   You will see a table under a header that reads, **Use Megaputer secure online ordering!**

6. In the table, go to the **TextAnalyst 2.1** row and click the appropriate checkboxes in the two rightmost columns.

7. Click **Finalize order** and follow the prompts.
   You will provide your registration number from step 3 of this procedure. You will also receive an activation code from Megaputer.

8. Click **File | Register**.
   The **Registration** dialog appears.

9. Paste your activation code in the bottom dialog box.
   You are now ready to use your TextAnalyst 2.1 software.

# Chapter 4: The TextAnalyst Interface

This chapter covers:

- Basic Navigation

- Accepted File Formats

- Knowledge Base Organization

- The TextAnalyst Workspace

Basic Navigation

This section discusses features and concepts that can help you use TextAnalyst 2.1 effectively. The TextAnalyst 2.1 user interface did not change significantly since the previous release.

The TextAnalyst Start Menu Items

During installation, TextAnalyst adds a TextAnalyst 2.1 program group to your Start menu. The group contains four icons.
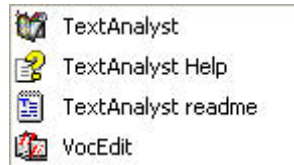
From top to bottom, these icons represent:

- **TextAnalyst** – Launches TextAnalyst, allowing you to create and open projects.

- **TextAnalyst Help** – Provides access to online help files that assist you in using TextAnalyst.

- **TextAnalyst Readme** – Provides version and registration information in the form of README.TXT file release notes.

- **VocEdit** – Launches the Dictionary Editor, allowing you to modify an existing dictionary or import your own custom dictionaries. For more information on VocEdit, see Chapter 8, Using Dictionary Editor.

The Startup Window

To start TextAnalyst 2.1, double-click the TextAnalyst icon on your desktop. You might also do the following:

1. Click the Windows **Start** button.
2. Choose **Programs | TextAnalyst 2.1,** and click the **TextAnalyst** icon or name. The TextAnalyst **Startup window** appears.
   If TextAnalyst 2.1 does not appear on your Windows **Start** menu, then:
3. Open Windows **Explorer**.
4. Navigate to **C:\PROGRAM FILES\MEGAPUTER INTELLIGENCE\MICROSYSTEMS\TEXTANALYST 2.1**
5. Double-click **TextAnalyst.exe**.
6. You have now started TextAnalyst and are ready to begin.

The Startup window provides three shortcut selections. You can:

- **Analyze new text and create a knowledge base:** Clicking this icon lets you have TextAnalyst analyze a new text, resulting in a knowledge base for that text.

- **Open an existing knowledge base:** Clicking this icon gives you access to a text and knowledge base you have already established in TextAnalyst.

- **View TextAnalyst tutorial:** Clicking this icon brings up the TextAnalyst Tutorial, which provides step-by-step instruction in using TextAnalyst.

If you choose not to use the Startup window, you can click the checkbox at the bottom of the window to prevent it from appearing on future startups.

Working with the Sample Text

It may be helpful for you to create a knowledge base from the sample text that TextAnalyst provides. This way, you can jump-start your work with objects in the workspace. To create a knowledge base from the sample, either select **Analyze new texts and create a knowledge base** in the shortcut Startup window, or follow these steps:

1. Launch TextAnalyst.
2. From the Main menu, choose File | Add new text…
3. Navigate to TextAnalyst 2.1/Examples and select Databasing in the 90s.txt. Click Open.

Accepted File Formats

TextAnalyst currently analyzes texts stored in the standard ASCII text file format (.TXT) and Rich Text Format (.RTF). Often you can convert text into .TXT or .RTF file formats by using a word processing application.

Additional File Formats for Future Releases

Future releases of TextAnalyst will include an expanded list of accepted file formats, including Microsoft Word (.DOC) documents, Adobe Acrobat files (.PDF), and HTML (.HTM) web pages.

The TextAnalyst Data (.TAD) File Format

For future reference, you can save the knowledge base TextAnalyst creates as a TextAnalyst Knowledge Base Data file (.TAD file format).

The knowledge base contains all results of semantic text exploration. This includes the complete semantic network of the investigated texts hyperlinked to corresponding sentences of original texts, as well as semantic weights of all important terms and the relationships between them.

Knowledge Base Organization

The Semantic Network

TextAnalyst automatically creates a semantic network when you load your text into TextAnalyst. This semantic network forms the foundation from which all other analyses stem.

We define **semantic weight** as the measure of the probability that a concept is contextually important. The organization of the semantic network is based on the semantic weight of concept. The semantic network lists concepts that are likely to have importance in the context of the original text along with the weight of the relationship between all the concepts in the text.

Associative Links

Concepts stored in the semantic network form hyperlinks to the sentences in which they occur in the semantic network. These sentences are also hyperlinked to the original sentences in the text you loaded. Together with the semantic network, the hyperlinks create easy navigation throughout the semantic network and your text. Hyperlinks facilitate other analyses, providing for analyses such as summarization and natural language querying of texts.

Topic Structure

TextAnalyst identifies the most important concepts from the semantic network and places them in a nested tree-like list. Topic Structure provides the hierarchy of concepts in your text in descending order with branches for subtopics.

You can invoke Topic Structure through the **Main** menu by clicking **View | Topic Structure**.

The TextAnalyst Workspace

Workspace Divisions

TextAnalyst's workspace is divided into three panes. These are the View pane, the Results pane, and the Text pane.
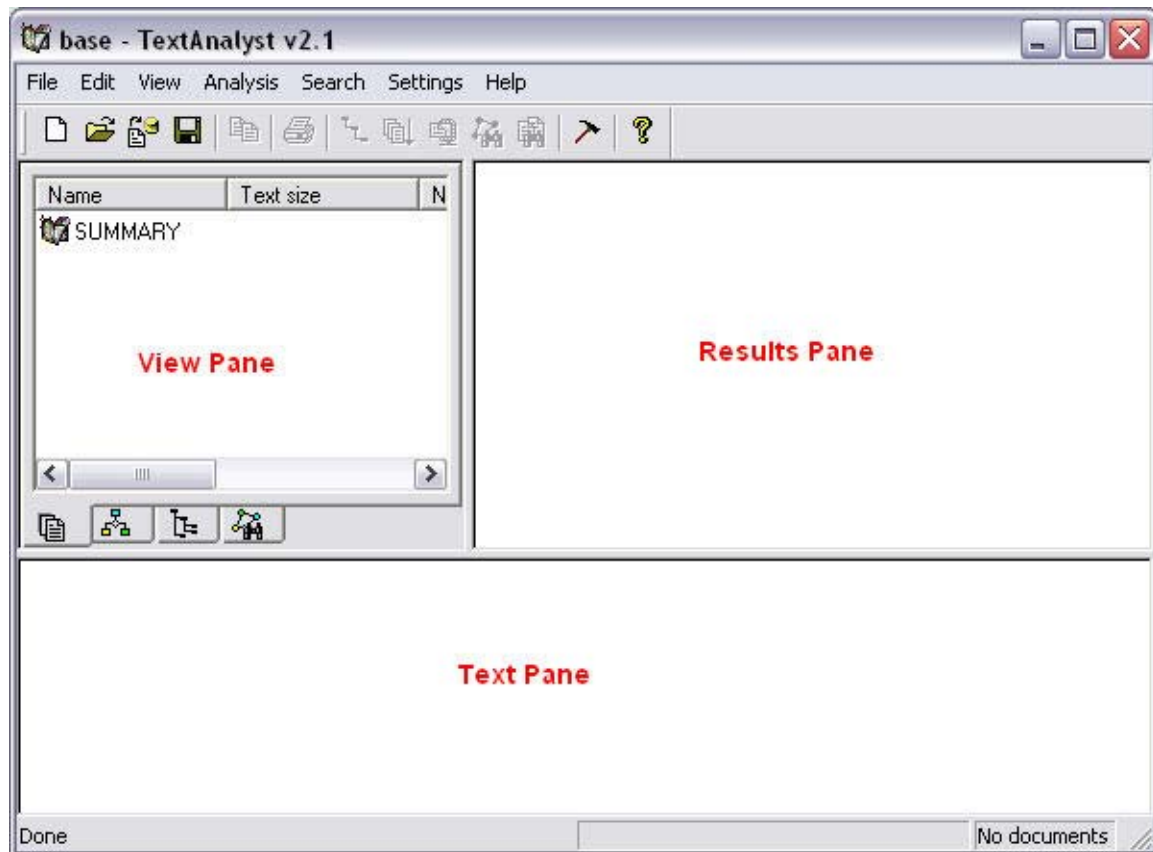
View Pane

The View pane is the upper left section of the workspace. This pane allows you to view five different types of displays. You can access these from the five tabs along the bottom of the pane, or from the View dropdown list in the Main menu. The displays, which we discuss in the next section, include:

- List of Documents
- Topic Structure
- Semantic Network
- Semantic Search Results

Results Pane

The Results pane is the upper right section of the workspace. This pane displays the results of TextAnalyst processing, such as a navigation, summarization, or semantic search. TextAnalyst shows every sentence containing the concept being examined, and highlights each occurrence of the concept in red.

Each sentence in the Results pane is hyperlinked to the original sentence in the Text pane. To go to the paragraph or sentence where a concept appears in the original text, double-click the related concept in the Results pane. The original text appears in the Text pane, showing the sentence you seek highlighted.

If you are working with more than one text, the text containing the sentence you seek appears in the Text pane, showing the sentence highlighted.

Text Pane

The Text pane is the bottom section of the workspace. This pane displays the currently analyzed text in its entirety. In addition, each concept in the semantic network is highlighted in green and underlined.

If you work with multiple texts, each text appears in the Text pane when you select it in the first tab of the View pane, List of Documents.

View Pane Tabs

Each View pane tab gives you a different way to view your analysis data. You choose the tab that gives you the information you need. The tabs are organized from left to right at the bottom of the View pane, and correspond from top to bottom on the View menu. The View menu is one of the Main menu items along the top of the workspace. Here are descriptions of each tab.

List of Documents Tab

This is first (leftmost) tab on the View pane and the first (top) selection on the View menu.
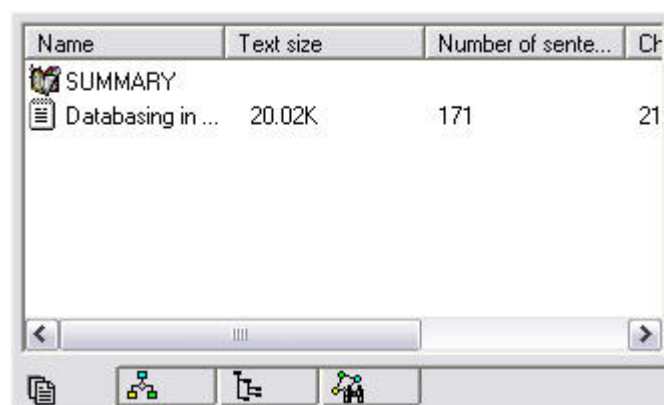
By selecting this tab, you can view a list of all the documents you have loaded for analysis. The summary is listed as a separate document. This summary appears when TextAnalyst generates the semantic network.

For each document, this tab tells you the text size, the total number of sentences the text contains, and the last date the text modified. The number in parentheses after the text size of the summary indicates the size of the summary relative to the size of the text(s) you are summarizing.
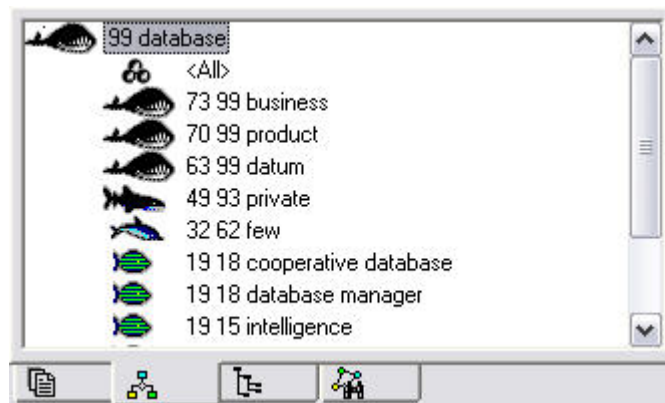
Topic Structure Tab

This is second tab on the View pane and the second selection on the View menu.

The topic structure reorganizes the previous tab (semantic network tab) by listing each concept individually with all its related concepts beneath it in a tree structure. Again, each concept is preceded by a pair of numbers. The first pair is the semantic weight of the relationship between a concept and its parent. The second is the weight of the individual concept alone.

Only one or a few items appear in the View window when you click the Topic Structure tab. You can double-click an item to expand it into the tree structure. To return to the more condensed representation, double-click the top item.

Semantic Network Tab

This is the third tab on the View pane and the third selection on the View menu.

The third tab displays the semantic network. The semantic network lists all the important concepts in your text and is the foundation for all further analyses TextAnalyst conducts. Words, word groups, and variations of single words are all significant in the semantic network.

Organization by Semantic Weight and Alphabetical Order
TextAnalyst calculates semantic weight using the statistical weights of individual concepts and the relationships between them. A concept having great significance in your document will have a high semantic weight.

By default, the semantic network shows concepts in the order of decreasing semantic weight. You can use the fish icons (the default) to help you assess relative semantic weights. The decreasing fish size helps convey decreasing semantic weights. To change the icons, see Chapter 7, Customizing TextAnalyst.

If you are searching for a concept and wish to see the semantic network in an alphabetized list, do the following:

1. From the **Main** menu, select **Settings | View Settings**.
   The Settings window appears.
2. Select the **Semantic network** tab.
3. Go to the first section, **Sort terms by:** Click the **alphabet** radio button.

Paired Numbers
There is a pair of numbers in front of each concept in the semantic network, with the exception of the parent concepts (concepts on the first level in the tree structure). Each number has a maximum value of 100.
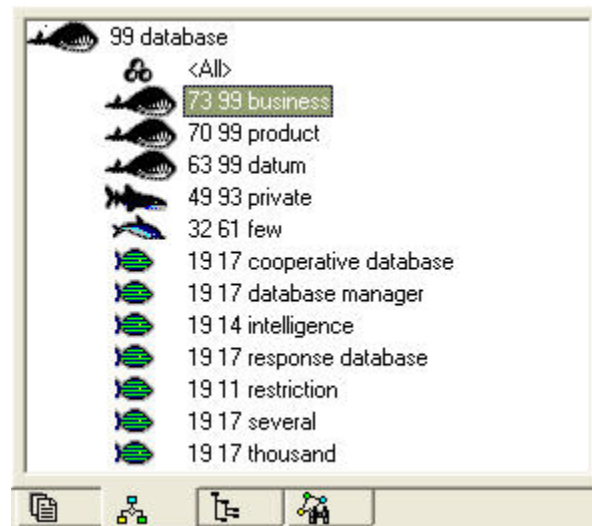
The first number represents the semantic weight of the relationship between a concept and its parent. For example, the concept "business" and its parent "databases," have a relational semantic weight of 73.

The second number indicates the semantic weight of the concept itself. For example, the semantic weight of "business" is 99.

Expanding and Collapsing the Tree Structure
You can click a plus sign in a square to the left of a term to expand the corresponding node of the tree structure. Clicking the plus sign reveals all terms semantically related to the term in the context of the investigated text.

Each related term has two numbers associated with it. The number on the left measures the semantic weight of the relationship of the new (daughter) term to the originally considered (parent) term. The parent term had the plus sign you clicked. The number on the right, located next to the term itself, represents the semantic weight of this term in the text.

Semantic Search Results/ Text Search Results Tab

This is the final and fourth tab on the View pane and the fourth selection on the View menu. Displays the semantic search and text search results.
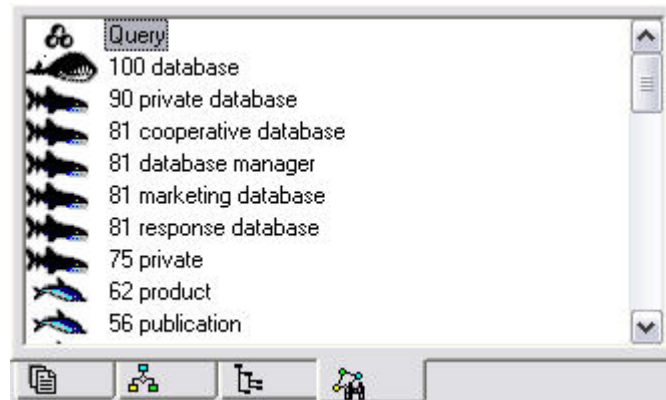
The Semantic Search Results tab displays the results when you conduct a semantic search, or natural language query. The results appear in a list of concepts related to your query. The list educates you about the context of the answer to your search. The list displays the semantic neighborhood of your answer, and provides you with additional words to use if you wish to make another query.

You conduct a semantic search from the Search menu item. For more information on using this item, see Using Semantic Search in chapter 5.

TextAnalyst's semantic search is unique and powerful because only TextAnalyst lets you type a question in conversational English and returns results based on ideas in your query instead of results that merely focus on keywords.

Text Search Results

Displays textual search results in the forth tab pane.
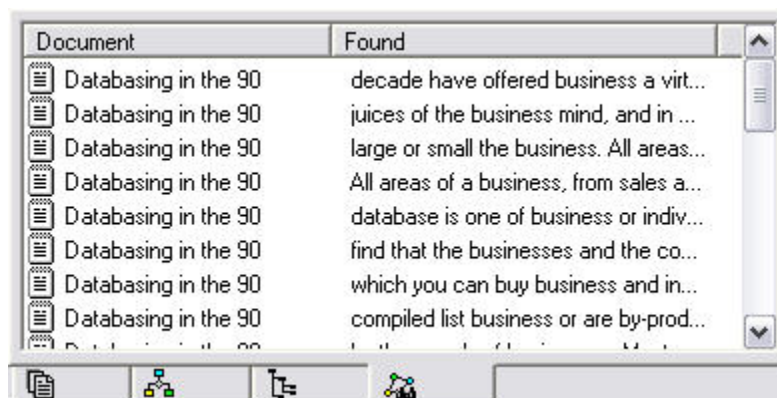
The Search Results tab is where you find the results of a standard search. You conduct this search from the Search menu item.

This tab displays every match, including the document name, or names if you search several texts, and a brief excerpt from the original text. The display helps you sift through the results quickly.

The results in the Search tab are hyperlinked to their location in the original text. To navigate to the search item, left-mouse click the icon, document name, or the excerpt. The search item then appears highlighted in the Text pane.

Main Menu

This section explains the function of each Main menu dropdown list. The Main menu extends along the top of the TextAnalyst workspace.

This section also provides the corresponding toolbar icon for a menu item, if the icon exists. Otherwise, the item simply appears with a round bullet in front of it. You can locate the toolbar just below the Main menu, if you have **View | ToolBar** checked. Clicking a toolbar icon gives you the same results as clicking the menu item.

If a menu item is shaded when you are using TextAnalyst, this means it does not apply to your current selections.

File Menu

Click the File dropdown menu for the following options:

Create: Allows you to create a new knowledge base. You will see a prompt to save the current knowledge base before closing it.

Open… : Allows you to open a knowledge base you previously saved in .TAD file format.

Add new text…: Allows you to locate and load a new text.

Save: Saves your current file.

- **Save as… :** Lets you rename your current file, saves it under that name, and lets you continue working.

- **Save results… :** Allows you to save the results of an analysis in a text file.

- **Export… :** Exports the semantic network to a .CSV file format for Microsoft Excel.

- **Export to HTML… :** Exports your file to the .HTML file format for viewing in a Web browser.

- **Print… :** This prints the contents of the current view in the Results pane.

- **Page preview… :** Lets you preview your document before printing.

- **Page setup… :** Lets you set margins for your printed page.

- **Register:** Lets you register TextAnalyst. To do this, you need an activation code from Megaputer. Registration allows you to use all the TextAnalyst features and to have eligibility for support.

- **Exit:** Allows you to save your work, then shuts down the software.

The Edit Menu

Click the Edit dropdown menu for the following options:

- **Copy:** Copies your highlighted text to the Windows clipboard.

- **Select all:** Selects the entire document in which you have single-clicked.

- **Rename:** Lets you renames a text when you select it in the View pane List of Documents. Select the text, then click **Edit | Rename**.

- **Delete:** Deletes a text from TextAnalyst after you select the text in the View pane List of Documents (first tab of the View pane). Select the text, then click **Edit | Delete**.

- **Break link:** Removes the link between two terms in the semantic network. You select a daughter term of a parent term in the Semantic Network tab or Topic Structure tab, then select **Edit | Break Link**.

- **Restore:** Allows you to restore changes to the semantic network or topic structure, including deletions, link breakages, and other changes.

The View Menu

The first five options on the View dropdown menu appear in detail under View Pane Tabs earlier in this section. The remaining options are:

- **Include all parents:** When checked, allows you to view all the parents of a single concept when you select that concept in the Semantic Network tab or the Topic Structure tab in the View pane. The default view shows one parent and all the subordinate concepts.

- **ToolBar:** When checked, allows you to access the toolbar icons under the Main menu.

- **StatusBar:** When checked, allows you to view the status bar along the bottom of the workspace. The status bar gives you data about your current document.

The Analysis Menu

Click the Analysis dropdown menu for the following options:

- **Content Analysis:** Performs the same semantic analysis that occurs when you load text into TextAnalyst. This feature lets you re-establish the semantic network after changing the dictionary linked to your TextAnalyst project.

- **Summarization:** Allows you to create a summary of your text having the length of your choice.

The Search Menu

Click the Search dropdown menu for the following options:

- **Semantic search:** Allows you to type a question in conversational English and returns results based on ideas in your query. Semantic search results appear in the Semantic Search tab on the View pane. For more information on semantic search, see Chapter 5.

- **Search:** Allows you to conduct a standard search for a word or phrase. Search results appear in the Search tab on the View pane. For more information on search, see Chapter 5.

The Settings Menu

Click the Settings dropdown menu for the following options:

- **General settings:** Brings up the General Settings window. The tabs in this window let you:

  - o Adjust routine startup displays
  - o Make changes to analysis conventions and access dictionary entries
  - o Change tree, window, and font settings

  For more information about using the General Settings window, see Chapter 7 of this manual.

- **View settings:** Brings up the View Settings window. The tabs in this window let you:

  - o Adjust the semantic network and the topic structure by changing the sorting of the root and the branches, and by changing the number of terms and links that appear.

  - o Adjust the number of sentences in the summary via a semantic weight requirement.

  - o Change the breadth of search results.

  - o Specify the semantic weight of hypertext terms.

- **Edit dictionaries… :** Brings up the Vocabulary Editor so you can view and edit several categories of words TextAnalyst handles for you. For detailed information on using this editor, see Chapter 8.

The Help Menu

Click the Help dropdown menu for the following options:

- **Contents:** Displays the table of contents for the online manual, which provides and index search capabilities.

- **TextAnalyst tutorial:** Opens the TextAnalyst online lessons, which give you hands-on lessons in using this product.

- **About the program**: Displays information about version 2.1 of TextAnalyst.

Chapter 5: Working with a Single Text

This chapter covers:

- Creating a Knowledge Base

- Understanding the Text Analysis

- Using the Tabs in the View Pane

- Navigating the Knowledge Base

- Editing the Knowledge Base

- Summarizing the Text

- Using Semantic Search

- Using Search

- Exporting to .CSV and .HTML File Formats

Creating a Knowledge Base

When you load a text, TextAnalyst automatically performs a semantic analysis of the text. A knowledge base results, consisting of:

- Your original text

- A semantic network

- Results of the analysis

- Hyperlinks between results and text

The knowledge base appears whether you are loading a new text, or opening a knowledge base you have saved before.

Loading Text

To load your text in TextAnalyst, follow these steps:

1. Launch TextAnalyst by following the steps under Basic Navigation in Chapter 4.

2. Go to the **Main** menu and either click **File | Open** for an existing text, or click **File | Add New Text…** for a text you have not loaded before.

3. Select the directory and the text you wish to load, then click **Open**.
   TextAnalyst begins working immediately. The knowledge base is complete once the **Please, wait…** separation of terms window disappears from your screen.
   This window displays the progress of the analysis. The top progress bar refers to the analysis of a single text, while the bottom progress bar applies to overall progress of all documents analyzed, whether one or more than one.
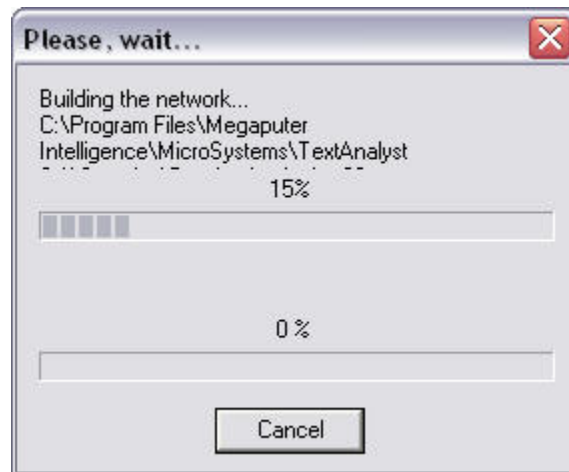
Loading the Sample Text

You might wish to practice using TextAnalyst before beginning an actual project. TextAnalyst provides a sample text for you. Follow the steps outlined in the previous section under Loading Text, but be sure to select the **TEXTANALYST 2.1/EXAMPLES** directory and the **DATABASING IN THE 90S.TXT** file.

Understanding the Text Analysis

Once you load a text, TextAnalyst immediately performs linguistic processing and semantic text analysis, resulting in the semantic network. TextAnalyst lets you examine information about your text in several ways. The information emerges from this semantic network.

You see the most basic representation of the semantic network from the **Semantic Network** tab (second tab from left) on the **View** pane. You can also see this tab by selecting **View | Semantic network** from the **Main** menu.

To understand the analysis and the various ways to look at the resulting data, become familiar with the section, "The TextAnalyst Workspace," in chapter 4. This section explains how the semantic network is organized and what the numbers in the tree structure represent, as well as different ways to view the analysis.

Using the Tabs on the View Pane

The View pane is the upper left section of the TextAnalyst workspace. There are five tabs in the View pane, each allowing you to view information in a different way. The tabs appear along the bottom of the View pane.

The Tabs

From left to right, these View pane tabs are:

- List of Documents tab
- Topic Structure tab
- Semantic Network tab
- Semantic Search Results tab
- Search Results tab

You can also access the tabs from the View menu on the Main menu bar.

For specific details about what you can do with each tab and how to access it, see View Pane Tabs under The TextAnalyst Workspace in Chapter 4.

Quick Reference

The following quick-reference table for the View pane tabs can help you use TextAnalyst.

If you want to:	Then click this tab/button:
See a list of your TextAnalyst documents	List of Documents – first tab
See the size of a summary	List of Documents – first tab
View a concept and all its related concepts	Topic Structure – second tab
See the tree structure of the knowledge base for a document	Semantic Network – third tab
Locate a specific term	Semantic Network – third tab
See the results of a semantic search	Semantic Search – fourth tab
Rephrase a semantic search	Semantic Search – menu button
See the results of a text search	Search – fourth tab

Navigating the Knowledge Base

As you use the knowledge base and the different tabs on the View pane, you might wish to locate specific text in your original document to read the context of a semantic concept. This section tells you how to find, in your original text, words and concepts that appear in the View or Results pane.

Every sentence of significant semantic weight is always linked to its original location, so you can easily navigate through the entire document.

Navigating from the Results Pane

Each sentence in the Results pane is hyperlinked to the original sentence in the Text pane. To go to the sentence where a concept appears in the original text, follow these steps:

1. In the **Results** pane, double-click the word or sentence you wish to find.

2. Look at the highlighted sentence in the **Text** pane.
   This is the entire original sentence containing the word or concept you seek.

Navigating from the View Pane

To find a word or concept you see in a tree structure on a View pane tab, do the following:

1. Click the word or concept in the **View** pane.

2. Locate the word or concept in the **Results** pane.
   This word is highlighted. Red is the default color.

3. In the **Results** pane, double-click the word or sentence you wish to find.

4. Look at the highlighted sentence in the **Text** pane.
   This is the entire original sentence containing the word or concept you seek.

Editing the Knowledge Base

You can edit the knowledge base by using the Edit menu to rename, delete, and break links between documents. You can also restore any changes before saving them.

Another way to modify the knowledge base is by making changes to the dictionary. For more information on customizing the dictionary, see Chapter 8.

Summarizing the Text

Summarization lets you quickly distill the meaning of a large text or collection of texts. This section reviews the process of creating a summary, then discusses how to interpret a summary and change its scope, or threshold.

Creating a Summary

To create a summary, follow these steps:

1. Load the document you wish to summarize.
   For assistance, see Loading Text under "Creating a Knowledge Base" in this chapter.

2. In the **View** pane **List of Documents** tab, highlight the document you wish to summarize.
   You can also see this tab by clicking **View | List of Documents** from the **Main** menu.

3. From the **Main** menu, select **Analysis | Summarization**.
   TextAnalyst processes the summary, then lists the sentences that make up the summary in the **Results** pane.

Interpreting the Results of a Summary

During summarization, TextAnalyst determines the semantic weight of each sentence and displays only sentences with a semantic weight higher than the threshold. TextAnalyst calculates the semantic weight of a sentence based on semantic weights of concepts in the sentence and the relationships between them.

The default threshold is 90. This means that all the sentences with a semantic weight greater than 90 appear in the Results pane.

The summary lists sentences with concepts that most likely have meaning in the context of the original text, when combined with the weight of the relationship between all concepts in the text.

The sentences included in a summary appear in the Results pane. These are the sentences with a semantic weight greater than the threshold. The sentences appear in the same order as in the original text. To locate the same sentence you see in the summary in your original text, click that sentence and it will appear highlighted in its context in the Text pane.

The View pane provides statistical information about your summary in the **List of Documents** tab. When you highlight the name of your text, the SUMMARY document at the top of the list tells you the:

- Size of the summary

- Percentage in size the summary is in comparison to the size of your original document

- Number of sentences the summary contains

Changing the Settings for a Summary

You can change both the semantic weight threshold and semantic weight display by doing the following:

1. From the **Main** menu, click **Settings**.
   The **Settings** window appears.

2. Click the **Summary** tab on the **Settings** window.
   The **Summary** tab appears, allowing you to change the semantic weight threshold and semantic weight display.
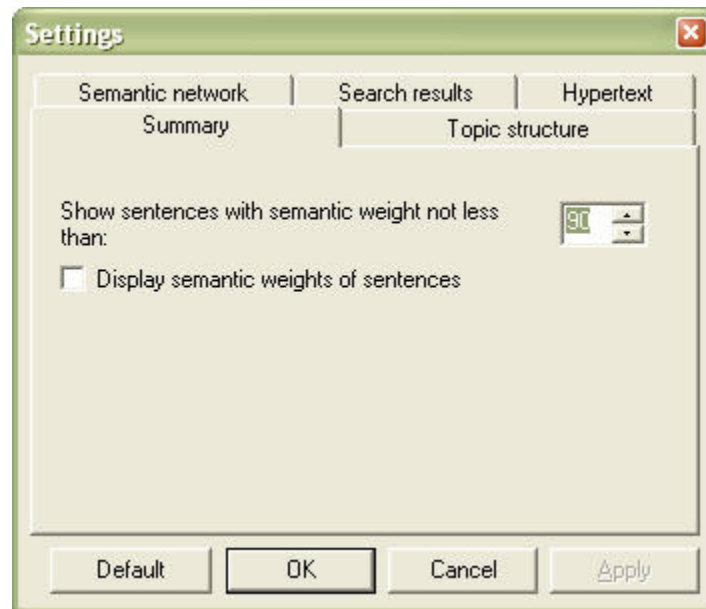
Changing the Semantic Weight Threshold

The semantic weight threshold is the number that indicates the minimum semantic weight of sentences TextAnalyst includes in your summary. The default is 90, so for any summary with the default threshold, all sentences with a semantic weight of 90 to 100 are included, 100 being the maximum weight.

Changing the semantic weight threshold increases or decreases the size of your summary. The lower the threshold, the more sentences TextAnalyst includes. The higher the threshold is, the fewer the number of sentences.

To change the semantic weight threshold, do the following:

1. Go to the **Summary** tab as directed in the start of this section.

2. Adjust the semantic weight threshold by using the arrow buttons or typing the number you wish to use.

3. Repeat your summary, observing the change in size.

Changing Semantic Weight Display

When you display the semantic weights of sentences, the numbers indicating semantic weight appear immediately in front of each sentence.

You can view or hide the semantic weight of each sentence in your summary in the **Results** pane. To do this, click the checkbox in the **Summary** tab to either check (view) or uncheck (hide) semantic weight display.

Printing a Summary

To print a summary, do the following:

1. Click anywhere in the Result pane to be sure the pane is active.

2. From the **Main** menu, click **File | Print**.
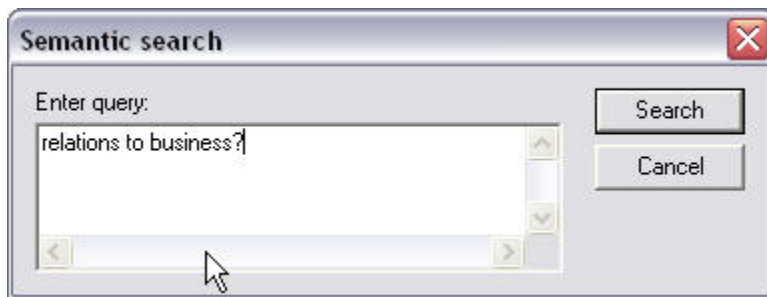
Using Semantic Search

Semantic Search provides the most powerful and search capabilities in TextAnalyst. Semantic search is synonymous with Natural Language Query. You make your queries using natural language, typing in a question in conversational English.

You use semantic search when you are asking a question or searching for ideas. You use standard search, which we cover in the following section, when you wish to locate a specific word or phrase in your text.

Starting a Semantic Search

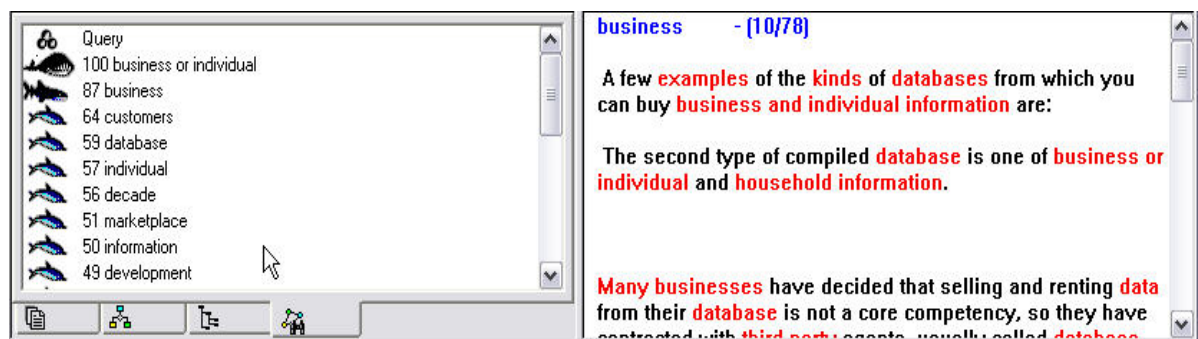To start a semantic search, do the following:

1. From the **Main** menu, click **Search | Semantic search…**
   The **Semantic Search** dialog appears.

2. Type your question, and then click **Search.**
   TextAnalyst searches your document.

Understanding Semantic Search Results

Semantic search results have two parts: The semantic structure in the view pane, and actual text in the View pane.

Semantic Search Results in the Results Pane

TextAnalyst pulls sentences from the original text that best respond to your query and places them in the Results pane. Each sentence is hyperlinked to its location in the original text.

Semantic Search Results in the View Pane

In the View pane, TextAnalyst provides a topic structure based on your query. To see this topic structure, either click the **Topic Structure** tab (second tab) in the **View** pane, or click **View | Topic structure** from the **Main** menu.

The Topic Structure tab can help if you are not sure how to phrase your question, or if the results to your first query do not yield meaningful results. The concepts relating to your query appear in the customary topic structure format.

TextAnalyst provides the expandable and collapsible tree structure so you can navigate the topic structure quickly. After viewing the topic structure, you may wish to rephrase your query for another semantic search that can better pinpoint the answer you seek.

You can adjust the number of sentences appearing in response to a query by clicking **Settings | View settings**, and then selecting the **Topic Structure** tab. You can broaden or restrict the results to obtain a large number of sentences or only the most relevant ones.

Using Search

TextAnalyst 2.1 includes a standard search utility similar to the "find" utility in many desktop applications. This utility lets you search for exact words and word combinations within your original text.

Conducting a Search

To search your text, do the following:

1. Click **Search | Search…** from the **Main** menu.
   The Search dialog appears.

2. Type the word you wish to locate, and then click **Search**.
   TextAnalyst locates your word in your original text.

The Search dialog provides a checkbox for **Only complete word** so you can restrict results to whole words that exactly match your query. Click this checkbox if you are looking for an exact word, and leave it unchecked if you also wish to see words containing what you typed.

For example, if you search for the word "phone" and do not click the checkbox, the results will include the words "phone" and "telephone". Clicking the checkbox means the word "telephone" would not appear as a match.

The checkbox next to **Utilizing the register** is not active for this release.

Using Search Results

You locate the results of your search in the View pane **Search** tab. Either click the **Search** tab (rightmost tab) in the **View** pane, or click **View | Search results** from the **Main** menu. The Search tab automatically displays the results of the query.

Every occurrence of the word you indicated appears in context to help you isolate the exact text for which you are searching. To find the word in your text, click the individual results listing. TextAnalyst highlights the sentence containing your word.

Exporting to .CSV and .HTML File Formats

You can export TextAnalyst's semantic network to a .CSV or .HTML file format.

Exporting to Excel (.CSV File Format)

To export a semantic network to a .CSV file format for use in Excel, do the following:

1. From the Main menu, click **File | Export…**

2. In the **Save As** window, indicate a filename and a directory.
    The file will have a .CSV file extension.

3. In the Microsoft Excel **Main** menu, click **File | Open**, and select the file you
    indicated in step 2.
    You semantic network appears, organized in columns.

The four columns in your Excel document are Parent, Frequency, Weight, and
Subordinate. Parent refers to the parent term in the tree structure, and Subordinate to the
subordinate term.

If the Subordinate column in Excel is blank, the numbers in the Frequency and Weight
columns pertain to the Parent column. If the Subordinate column is not blank, numbers in
the Frequency and Weight columns pertain to the relationship between the parent and
subordinate.

Exporting to the .HTML File Format

To export a semantic network to a .CSV file format for use in Excel, do the following:

1. From the Main menu, click **File | Export to HTML…**

2. In the **Save As** window, indicate a filename and a directory.
    The file will have an .HTML file extension.

3. Go to your Web browser, click **File | Open**, and select the file you indicated in step 2.
    You semantic network appears with hyperlinks for navigation.

Clicking a concept in your .HTML file shows that concept in the context of the Results
pane. Clicking the concept again from those results shows its occurrence in the original
text. To return to the semantic network, click the hyperlinked header term.

Chapter 6: Working with a Text Base

This chapter covers:

- Creating a Knowledge Base from Several Texts
- Analyzing Several Texts

Creating a Knowledge Base from Several Texts

This chapter covers work you do with a group of related texts in TextAnalyst. When you create a knowledge base of such texts, you select these texts at one time and TextAnalyst processes them in order, producing a single summary for all the texts.

Creating the Base

To place a group of related texts in a knowledge base, follow these steps:

1. Place all the texts in a single directory on your computer.

2. From the TextAnalyst **Main** menu, click **File | Open**.

3. Navigate to the directory where you placed your texts.

4. Select all the texts you wish to include in the knowledge base.
   You **Ctrl – Click** to select individual texts, or **Shift – Click** to select groups of texts.

5. Click open.
   TextAnalyst creates your knowledge base for the texts you selected, and includes a single summary of all the texts.

Adding More Texts to the Base

You add more texts to an existing knowledge base by doing the following:

1. Close the knowledge base.

2. Create a new knowledge base as outlined above, including all the texts from the previous base as well as any new texts you wish to add.

Analyzing Several Texts

Analysis of a base of several texts is similar to analysis of a single text. A few differences are:

- **Progress Bar** - The second (bottom) progress bar appearing during analysis shows you the progress of all the documents in your text base. The first (top) progress bar shows you the progress of the current document.

- **Navigation** - When you navigate using hyperlinks, clicking in the View and Results panes means you are selecting concepts found among all the documents. Locating information in the Text pane, however, means you are identifying a sentence from one of the texts in your base.

- **Clustering** - Clustering occurs automatically when you create your knowledge base. In the Topic Structure tab, you can find concepts from your entire knowledge base grouped in the tree structure according to the relationships between them. To find clusters of related information across several documents, click the parent concept of a cluster, then locate and click the hyperlinks in the Results pane.

Chapter 7: Customizing TextAnalyst

This chapter covers:

- General Settings
- View Settings

General Settings

To locate the **General Settings** dialog, click **Settings | General settings**. Three tabs are visible.

Startup Tab

The first two checkboxes are important.

The **Do not display startup dialog box** checkbox lets you control whether or not the startup shortcut window appears.

The **Summarize analyzed texts automatically** checkbox lets you indicate if you want TextAnalyst to generate a summary when loading your text.

The **Analyze metainformation** checkbox should be checked in order to correctly identify situations when a period does not represent the end of the sentence. This feature is important when processing text with name initials, abbreviations, internet addresses, etc.

When the **Store supplementary information** checkbox is checked, TextAnalyst stores on the hard drive an additional file with the extension "mpp" for each processed text. This file contains an image of the text after the initial linguistic preprocessing. Storing the results of linguistic preprocessing significantly speeds up the work with large textbases. Linguistic preprocessing might be the most time consuming step in the analysis of a text. When the user ads a new text to the existing large TextAnalyst knowledge base, the whole knowledge base has to be recalculated to take into account the presence of new information. In this case, the stored results of linguistic preprocessing of each text help significantly increase the speed of analysis, there is no need to repeat time consuming linguistic preprocessing on these texts.

Analysis Tab

For the first two checkboxes:

Analyze word combinations lets you include concepts that have more than one word, or, when checked, see only one-word concepts.

When checked, **Analyze only unconditionally included words** tells TextAnalyst to analyze only those words you specify as unconditionally included in the dictionary and their combinations with other words.

View Tab

In the View tab upper left area, you can select a fish icon or a standard tree indicator. Numerical values in the tree structure can either be according to semantic weight (recommended) or frequency, which is the number of times a concept appears in the text(s).

On the right side in the View tab, you can adjust settings for the View pane (number 2) and the Text pane (number 3). You can change color and font.

View Settings

To locate the **View Settings** dialog, click **Settings | View settings**. Five tabs are visible.

Topic Structure tab – This tab lets you adjust the topic structure by changing the sorting of the root and the branches, and by changing the number of terms and links that appear.

Summary tab - Adjusts the number of sentences in the summary via a semantic weight requirement.

Semantic Network tab - This tab lets you adjust the semantic network structure by changing the sorting of the root and the branches, and by changing the number of terms and links that appear.

When the **Hierarchy** check box is checked, the lower levels of the semantic network displays only those concepts, which have not been encountered in the higher levels.

Search Results tab – Lets you change the breadth of search results.

Hypertext tab – Allows you to specify the semantic weight of terms appearing in hypertext.

Chapter 8: Using Dictionary Editor

This chapter covers:

- Dictionary Basics

- How to Edit the Dictionary

Dictionary Basics

TextAnalyst uses a dictionary for each analysis. The basic dictionary that accompanies TextAnalyst (stored in a file *normal_eng.dic*) is a result of initial automated and manual training of the system on a very large corpus of texts from many different domains. The presupplied dictionary contains **common words**, **words not analyzed**, and **exception words**. Words from each of these three groups must be treated specially during text linguistic preprocessing and semantic analysis.

Important note: **All regular words, which are good candidates to become semantically important concepts in a text are not included in the dictionary.**

The dictionary can be manually edited and saved under a new name in order to fine-tune TextAnalyst to perform its best in a certain application domain.

The dictionary tells TextAnalyst:

- Which concepts, if any, are **user words**. These concepts are the ones you specifically want to be included in the semantic network whether TextAnalyst finds them to be semantically important in the context of the analyzed text or not.

- What the **common words** are. These are words that are assumed to have little semantic importance on their own and are used as modifiers with other semantically important words. Many adjectives, as well as some verbs and nouns are treated by TextAnalyst by default as common words. TextAnalyst does not analyze common words unless they are combined with another word, creating a semantically important concept.

- The concepts that are **not analyzed words**. These are articles and words you indicate that should be ignored by TextAnalyst.

- The **exception words**. These concepts are words that do not follow usual rules of stemming. Most often, exception words are represented by verbs with irregular verb forms.

As TextAnalyst analyzes your text, it works from both your text and the dictionary to calculate semantic weights. The dictionary helps TextAnalyst "know" which words to analyze.

You can use TextAnalyst without accessing the dictionary. However, editing the dictionary allows you to obtain more accurate analyses. The next section helps you edit the dictionary.

How to Edit the Dictionary

To access the dictionary for editing, click **Settings | Edit dictionaries…** from the **Main** menu. The **VocEdit** window appears.
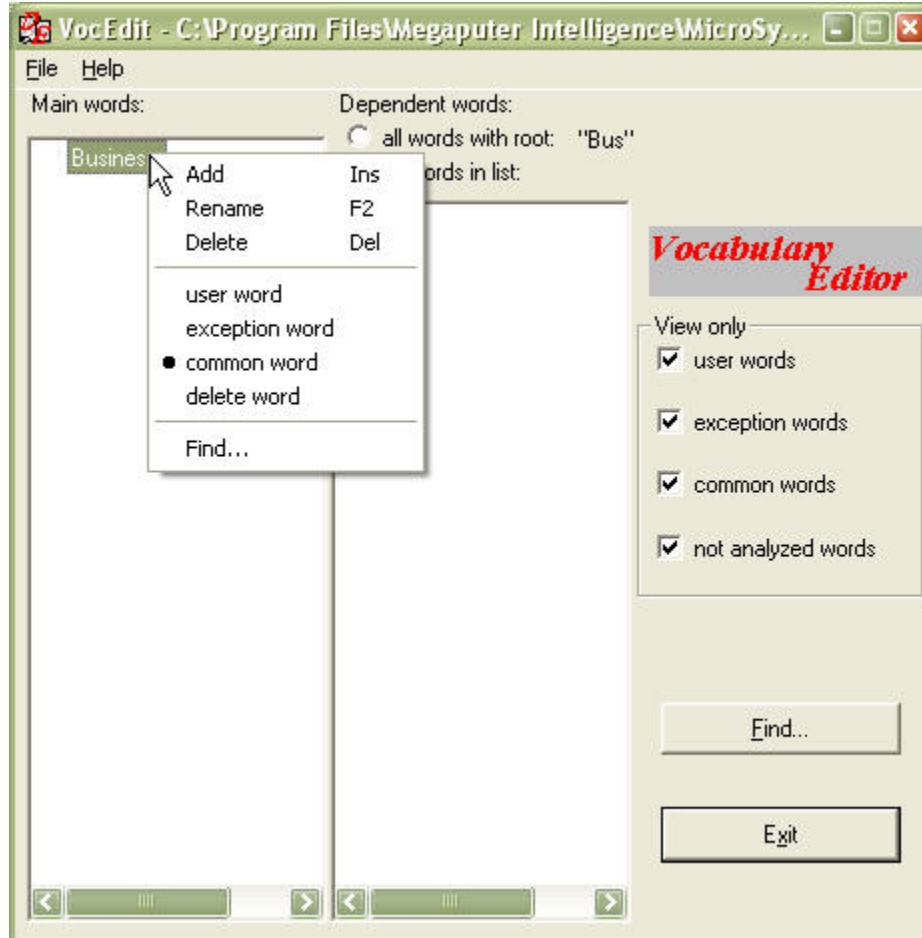
The left side of the VocEdit window shows **Main words**. Words appear in the Main words list only if you include them by using the **View only** checkboxes on the right side of the window. For example, you will view only the exception words if you check the second checkbox and leave the other boxes unchecked. To see the entire dictionary in the Main words box, check all the checkboxes under **View only**.

In the list of Main words, you may select their type:

- User words have a **red checkmark** to the left of them

- Exception words have a **green diamond** to the left of them

- Common words have a **blank space** to the left of them

- Not analyzed words have a **black "X"** to the left of them

The middle column of the dictionary window shows **Dependent words**. These words are the variations of root words. All the dependent words encountered in an analyzed text

will be treated as identical synonyms of the main word. Each dependent word encountered in the text will be substituted by the corresponding main word.

Modifying the Main Words List

To add a word to the list, right click with your cursor in the Main words column, then click add. Type your word. By default, this word will be a common word.

You click a word to highlight it. With your cursor on the highlighted word, right click to designate the word as a user, exception, or common word. Clicking delete in the middle section under common word marks the word with an "X" – not analyzed.

When you right click in the Main words list, the top three choices are add, rename, and delete. Choosing delete in this upper section deletes the word, while choosing delete in the middle section marks the word as not analyzed.

You can modify the Dependent words list in a way similar to the Main words list. Just right click with your cursor in the Dependent words field.

Saving the Dictionary

After you modify the dictionary, you can click **File | Save** from the dictionary **Main** menu to save it. TextAnalyst continues to use the same dictionary for all analyses unless you specifically link a different dictionary to be used with a certain knowledge base (.TAD file). To save an edited dictionary click **File | Save as** from the dictionary **Main** menu. TextAnalyst saves the dictionary as a .DIC file. You can link the new dictionary to a selected knowledge base by using **Settings | General Settings | Analysis | Dictionary** control and selecting the desired dictionary file from the appearing dialog box. The next time you open a text, TextAnalyst will use the new dictionary.

For easy identification, we recommend naming the .DIC file with a name similar to the .TAD file you wish it to accompany.

Loading User-prepared Dictionaries

In order to add a substantial number of words to TextAnalyst dictionary, the user can load utilized dictionaries prepared with a regular text editor such as Notepad.

The created dictionary file must be in .txt format and must follow the following specifications:

MainWord1*/A1/A2 CR*
Synonym11 *CR*
Synonym12 *CR*
...
Synonym1N *CR*
***** CR

MainWord2*/A1/A2 CR*
Synonym21 *CR*
Synonym22 *CR*
...
Synonym2K *CR*
***** CR

...

***** CR

MainWordM*/A1/A2 CR*
SynonymM1 *CR*
SynonymM2 *CR*
...
SynonymML *CR*
*****

Here *CR* stays for the Enter key (Carriage Return), no blank spaces should be present, and the modifiers *A1* and *A2* have the following meaning:

A1 = 10 -- All words in an analyzed text that have the same root as the corresponding **Main Word** are identified with this **Main Word**. In this case, the list of synonyms is either empty, or is ignored in the analysis.

A1 = 11 -- All words listed as synonyms are identified with the corresponding **Main Word**. The list of synonyms might be empty. This setup is used when the user wants to prevent words with the same root from being identified with the corresponding **Main Word** in the analysis.

A2 = 101 -- The corresponding Main Word is a **common word**

A2 = 102 -- The corresponding Main Word is an **exception word**

A2 = 103 -- The corresponding Main Word is a **user word**

A2 = 104 -- The corresponding Main Word is a **not analyzed word**

By default, **Main Words** with missing modifiers are assumed to be **common words**, and only the words listed as synonyms are identified with them.

Example: *Additional_dictionary.txt*

> Megaputer*/11/103 CR*
> ***** *CR*
>
> unify*/10/101 CR*
> ***** *CR*
>
> supervisor*/11/103 CR*
> lead *CR*
> manager *CR*
> leader *CR*
>
> ***** *CR*
>
> mistake*/11/102 CR*
> mistook *CR*
> mistaken *CR*
> mistaking *CR*
> mistakes *CR*
> *****

The user can either *Load* a new dictionary to be used by TextAnalyst in its analysis, or *Add* an extra list of words to an existing TextAnalyst dictionary.

In both cases, the user starts by selecting the **Settings | Edit Dictionaries** option from the Main menu of TextAnalyst.

In order to *load* a new dictionary prepared with a text editor as specified above, select **File | Open** option of the **VocEdit** pull-down menu, change the **File of type** setting of the opening dialog box to *.txt*, and then locate and open the prepared dictionary file on your machine. The file you have loaded substitutes the current dictionary that was used by TextAnalyst.

In order to *add* a new list of words to an existing TextAnalyst dictionary, select **File | Add vocabulary…** option of the **VocEdit** pull-down menu and then locate and open the prepared dictionary file on your machine. Words in the opened file are added to the current dictionary that is used by TextAnalyst.

> **Important note**: If an added dictionary contains words that coincide with words already present in the current dictionary, these words from the new dictionary are ignored.